An improved hybrid of SVM and SCAD for pathway analysis

Abstract:

Pathway analysis has lead to a new era in genomic research by providing further biological process information compared to traditional single gene analysis. Beside the advantage, pathway analysis provides some challenges to the researchers, one of which is the quality of pathway data itself. The pathway data usually defined from biological context free, when it comes to a specific biological context (e.g. lung cancer disease), typically only several genes within pathways are responsible for the corresponding cellular process. It also can be that some pathways may be included with uninformative genes or perhaps informative genes were excluded. Moreover, many algorithms in pathway analysis neglect these limitations by treating all the genes within pathways as significant. In previous study, a hybrid of support vector machines and smoothly clipped absolute deviation with groups-specific tuning parameters (gSVM-SCAD) was proposed in order to identify and select the informative genes before the pathway evaluation process. However, gSVM-SCAD had showed a limitation in terms of the performance of classification accuracy. In order to deal with this limitation, we made an enhancement to the tuning parameter method for gSVM-SCAD by applying the B-Type generalized approximate cross validation (BGACV). Experimental analyses using one simulated data and two gene expression data have shown that the proposed method obtains significant results in identifying biologically significant genes and pathways, and in classification accuracy.